This issue of the Bulletin *is devoted to some of the work now being undertaken in the area of standards. A number of initiatives underway relate to the description of machine readable records in inventories or union lists and the transfer of machine readable records from one system to another. These issues are of major concern from an archival point of view in that they help to provide data in a format for transfer and preservation, and they provide information about data holdings to a wide variety of research communities.*

CULDAT Record Format

As was indicated in the last issue of the *Bulletin*, the CULDAT (Canadian *Union* List of Machine Readable *Data* Files) is continuing. During the first phase of the project a draft record standard was developed. Since that time a revised record standard has been developed and discussed (IASSIST 1987). The following is a description of the standard which will be used for the descriptive entries. From the outset, a central consideration in the design of the content and form of the CULDAT MRDF description has been compatibility with existing standards for bibliographic control of computer files. The format has attempted to accommodate, as much as possible, these standards as well as the divergent needs of a variety of applications and products, like a data inventory.

The CULDAT descriptive standard has been limited to those data elements considered most useful in relation to the principal project objective — to establish a centralized source of information on computer files in Canada that will assist Canadian researchers in identifying and locating files of potential use. Data elements for MRDF description fall into six basic categories:

(1) Bibliographic elements, i.e., those needed to identify a MRDF.
(2) Descriptive elements to describe the contents of a MRDF.
(3) Classification elements necessary to group like data files together.
(4) Technical and physical elements that are needed to access a MRDF.
(5) Elements required to analyse or use a MRDF.
(6) Archiving elements to archive and maintain a MRDF.[1]

[1]Dodd, Sue A. *Cataloguing Machine-Readable Data Files: An Interpretive Manual.* Chicago: American Library Association, 1982. 247 p.

For the purposes of CULDAT the two primary categories are the bibliographic and descriptive data elements. Technical and physical information required for accessing computer files, and elements required to use and analyze a file, fall outside the scope of the national union list. This information will be provided by the appropriate data archive at the local level. An exception to this policy is the inclusion of a description of the study methodology, which is considered to be significant descriptive information in the early stages of determining whether a data file is of potential interest. The nature of a computerized inventory renders the general use of classificatory elements unnecessary — the bringing of like MRDFs together is accomplished through the search and retrieval process for each particular usage requirement. Several elements which are classificatory in nature have been included as discrete fields to enable easy partitioning of the data base; for example, the "Type of File" field enables the user to retrieve all numeric time series data files, and then search by other criteria.

A list of the CULDAT data elements, by category, is given below.

Bibliographic Identity

Principal Investigator(s): This field is used to record the name(s) of the person(s) chiefly responsible for the creation of the intellectual content of the data file.

Corporate Investigator(s): This field is used to record the name(s) of one or more corporate bodies chiefly responsible for the creation of the intellectual content of the data file, when there is no named principal investigator.

Title: This field is used to record the full title of the MRDF, including the subtitle, alternative title and/or parallel title (title translated in another language).

Variant Title: This field is used to record a title that is different from the title given in the Title field and which the cataloguer deems significant as additional identifying information or as an additional access point.

Edition Statement: This field is used to record information relating to an edition of a data file that contains differences from other editions.

Production: This field records information on the place of MRDF production, the name of the producer and the date of production. The

producer of MRDF is defined as the person or corporate body with the financial and/or administrative responsibility for the physical processes whereby the data file is brought into existence. Production date is defined as the date the file became operational in a machine readable form and available for analysis, processing and possible release to the public.

Distributor: The name of the distributor and the order number assigned by a distributor to the MRDF, as applicable, are recorded in this field. The distributor is that person or organization designated by the investigator or producer to reproduce copies of a particular file for distribution.

Generator: A generator is a person or corporate body with the responsibility for the physical processes by which a file is converted into machine readable form but whose responsibility does not extend beyond data conversion. The name of the generator is included in this field.

File Size: This field contains a statement of the number of parts (i.e., component files) that make up a MRDF, and the number of logical records in each file.

Series Statement: If the MRDF is part of a monographic series, this field records the title of the series and the volume or number of the part if the series is numbered.

Sponsor: This field is used to record information on the sponsor or funding agencies when the MRDF results from a funded project.

Descriptive Elements

Contents Note: This field is used to record the titles and size of individually named parts (files) of a MRDF in a formatted note. It is also used to give unformatted contents notes for serials specifying important special items with specific titles, unless they are catalogued separately. The number of logical records for each file is given following the title.

Frequency of Issue: This field is used only for MRDF issued serially to record data on the frequency of issue.

Abstract: The abstract contains a comprehensive description of the purpose, nature and scope of the file, special characteristics of the contents, major subject areas covered, and methodology, as applicable. The abstract includes two subfields, text and variable names.

Time Period: This field is used to record information on the time period to which the data contained in the file refer.

Number of Cases: This field gives the number of cases (units of analysis) in the file and the number of variables per case. The units of analysis are named.

Access Restrictions: This field is used to record information about restrictions imposed on access to the MRDF. A brief statement of any legal, physical or procedural restrictions, as imposed by the principal investigator or producer, is given here.

Classification Elements

Geographic Coverage: This field is used to record information about the geographic coverage of the MRDF, and contains the names of geographic areas, political jurisdictions and other geographic designations to which data apply.

Type of File: This field is used to record information which characterizes the file. In addition to a general descriptor (e.g., text, numeric) more specific information describing the nature of the contents, such as the form or genre of textual material (e.g., law reports) may be recorded in this field.

Access of MRDF

Software Dependency: If a file is dependent upon a particular data base management or statistical software system, this field contains a statement indicating this dependency.

The elements listed are only described briefly. A more detailed description of the elements can be obtained from the Government Archives Division of the National Archives.

ESRC Seminar Series on the Bibliographic Control of Computer Files

A series of meetings have been held in the United Kingdom in an attempt to bring together the views of representatives of various agencies and individuals from academic, government and private sectors who have a stake in the creation of a national standard for describing and cataloguing computer files. Among the various groups invited to participate were: data producers; information technologists and scientists; librarians; archivists; data and information disseminators; and data users.

The need for a coordinated national cataloguing effort seemed widely accepted. The formation of this seminar series is a response to the need for a coordinated program of action. The group has agreed to continue to meet with the following objectives:

(1) to standardize the cataloguing of computer files in line with the recommendations of the Anglo-American Cataloguing Code (2nd ed.) and the International Standard Bibliographic Description within the MARC format; (2) to seek funds to set up a pilot project to create a union catalogue of the computer readable holdings of the University of Edinburgh and the ESRC Data Archive, with descriptions of software; (3) to formulate and implement a policy of publicity for the project, including incentives and guidelines for producers, cataloguers and users of computer files and to investigate the possibility of setting up a clearing house for the deposit of cataloguing records of computer files.

To pursue these objectives, a number of working groups have been formed to study specific subjects.
(1) The Working Group on Standards is to investigate the adoption of a MARC record which would accommodate information required by producers and users of computer files in the United Kingdom and facilitate the exchange of information about computer files.
(2) The Working Group on Implementation is responsible to implement the creation of a trial union catalogue.
(3) The Working Group on Dissemination will formulate a policy for publicity and incentives and develop guidelines for participating organizations.

Funding has been obtained to continue the work of the seminar series and related working groups. Progress on the adoption of standards is being made and the work of this group is being followed with interest.

Bureau of Canadian Archivists and Descriptive Standards

Following up the recommendations of the report *Towards Descriptive Standards*, the Bureau of Canadian Archivists has formed a planning committee to oversee the establishment of descriptive standards and rules for the intellectual control of archival materials at four levels — the fonds d'archives, the series, the file and the item. The standards for description at the fonds d'archives level will include all material; the standards developed for the three other levels of description will be specific to the archival medium.

The mandate of the committee is to plan the development of descriptive standards based on accepted archival theory and principles for the Canadian archival profession. As well as the establishment of descriptive standards and rules for the four levels, the committee will work with the Canadian Committee on MARC and the National Library of Canada to adapt the format for archival material; study the existing name authorities and rules and adapt them to archival principles and practices; study the problems of subject indexing of archives and archival finding aids; publish and disseminate the work of the Planning Committee on Descriptive Standards; and study the mechanisms necessary for the ongoing maintenance and control of descriptive standards once developed.

The first working group to establish standards and rules for the intellectual control of archival material at the fonds d'archives level was appointed and has begun their work. This level of description will include all media, including machine readable records. The working group consists of five media specialists representing textual archives; maps and architectural archives; photographs, paintings, prints and drawings; machine readable archives; and sound and moving image materials. The work of the group is to be completed by March 1988.

Le Bureau canadien des archivistes et les normes de description

Conformément aux recommandations du rapport intitulé *Les normes de description en archivistique*, le Bureau canadien des archivistes a constitué un comité de planification chargé de surveiller la création de normes de description et de règles pour le contrôle intellectuel des documents d'archives de quatre niveaux : le fonds d'archives, la série, le dossier et le document. Les normes relatives au niveau du fonds d'archives porteront sur tous les types de documents; quant aux normes élaborées pour les trois autres niveaux, elles seront adaptées à chaque support archivistique.
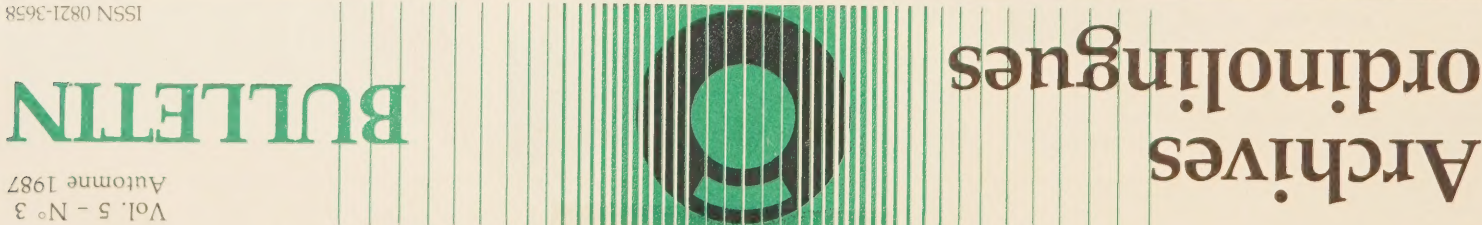
Le mandat du comité consiste à planifier l'élaboration de normes de description fondées sur les théories et les principes reconnus au Canada dans le domaine de l'archivistique. En plus d'élaborer des normes de description et des règles sur les quatre niveaux susmentionnés, le comité travaillera, en collaboration avec le Comité canadien du MARC et la Bibliothèque nationale du Canada, à adapter le format MARC pour les documents d'archives, à étudier les vedettes d'autorité et les règles existantes et à les rendre conformes aux principes et aux pratiques archivistiques, à étudier les problèmes liés à l'indexage des sujets des documents d'archives et aux instruments de recherche en archivistique, à publier et à diffuser les résultats des travaux du comité de planification des normes de description, et au contrôle des normes de description élaborées.

Le premier groupe de travail, chargé d'élaborer des normes et des règles au niveau du fonds d'archives, est déjà formé et il a commencé son travail. Ce groupe de travail se compose de cinq spécialistes dans les domaines suivants : les archives textuelles; les cartes et les archives de l'architecture; les photographies, les peintures, les estampes et les dessins; les archives informatiques; ainsi que les documents audiovisuels. Le groupe doit avoir terminé son travail d'ici mars 1988.

Les articles ont été rédigés par Sue Gavrel de la Division des archives gouvernementales.

Série de séminaires de l'ESRC sur le contrôle bibliographique des fichiers informatiques

On a tenu une série de réunions au Royaume-Uni pour tenter de concilier les vues des représentants de divers organismes et de particuliers provenant du secteur universitaire, gouvernemental et privé, qui souhaitent l'élaboration d'une norme nationale de description et de catalogage des fichiers informatiques. Parmi les groupes invités à participer aux discussions, il y avait des producteurs de données, des technologues de l'information, des documentalistes, des bibliothécaires, des archivistes, des diffuseurs de données et d'information, et des utilisateurs de données.

La nécessité d'activités de catalogage coordonnées à l'échelle nationale semble généralement reconnue. L'organisation de la série de séminaires en question répond au besoin de dresser un plan d'action coordonné. Les membres du groupe ont accepté de continuer à se réunir pour atteindre les objectifs suivants : 1) normaliser le catalogage des fichiers informatiques conformément aux recommandations des Règles de catalogage anglo-américaines (deuxième édition) et de la Description bibliographique internationale normalisée; en adoptant le format MARC; 2) tenter d'obtenir des fonds pour pouvoir mettre sur pied un projet pilote visant à créer un catalogue collectif des fonds ordinolingues de l'Université d'Edimbourg et des Archives de données informatiques de l'ESRC, contenant des descriptions de logiciels; 3) formuler et mettre en application une politique publicitaire et incitatrice et des lignes directrices à l'intention des producteurs, des catalogueurs et des utilisateurs de fichiers informatiques, et étudier la possibilité de créer un centre pour la conservation des notices catalographiques des fichiers informatiques.

Pour que ces objectifs soient atteints, plusieurs groupes de travail ont été constitués :

1) Le groupe de travail sur les normes doit étudier la possibilité d'adopter une notice MARC qui renfermerait l'information nécessaire aux producteurs et aux utilisateurs de fichiers informatiques du Royaume-Uni et qui faciliterait l'échange d'information concernant les fichiers informatiques.

2) Le groupe de travail sur la mise en application est responsable de la création d'un catalogue collectif provisoire.

3) Le groupe de travail sur la diffusion formulera une politique publicitaire et incitatrice, et élaborera des lignes directrices à l'intention des organismes participants. On a obtenu des fonds permettant la poursuite des travaux sur la série de séminaires et sur les groupes de travail connexes.

(fichiers) d'un fichier ordinolingue. Cette zone peut également contenir des notes non structurées concernant des publications en série et indiquant les articles spéciaux importants avec leur titre, à moins que ceux-ci ne soient catalogués séparément. Le nombre d'enregistrements logiques que comporte chaque fichier est fourni après le titre.

Cadence de parution : cette zone n'est utilisée que pour les fichiers ordinolingues publiés en série; on y note la fréquence de parution.

Résumé : cette zone renferme une description complète de l'objet, de la nature et de l'envergure du fichier, et de l'information sur les principaux domaines touchés et les caractéristiques spéciales du contenu, les utilisées, le cas échéant. Cette zone comporte deux sous-zones, du texte et le nom de variables.

Période : dans cette zone, on note la période visée par les données que le fichier renferme.

Nombre de cas : cette zone renferme de l'information sur le nombre de cas (unités d'analyse) du fichier et sur le nombre de variables par cas. Les unités d'analyse y sont nommées.

Restrictions concernant l'accès : on note brièvement dans cette zone toute restriction concernant l'accès au fichier ordinolingue. Imposées par l'enquêteur principal ou le producteur, les restrictions peuvent être d'ordre juridique, physique ou procédurel.

Éléments de classification

Champ géographique : cette zone renferme de l'information sur le champ géographique du fichier ordinolingue, ainsi que le nom des régions, des circonscriptions et des autres lieux géographiques auxquels les données s'appliquent.

Type de fichier : on note dans cette zone les caractéristiques du fichier. On peut y indiquer, en plus d'un descripteur général (exemples : texte, nombres), des renseignements plus précis sur la nature du contenu tels que la forme ou le genre de l'écrit (exemple : recueil de jurisprudence).

Consultation d'un fichier ordinolingue

Dépendance à l'égard d'un logiciel : si un fichier dépend de la gestion d'une base de données ou d'un système logiciel statistique en particulier, cette dépendance est indiquée dans cette zone.

Les éléments susmentionnés ne sont décrits ici que brièvement. On peut en obtenir une description plus détaillée auprès de la Division des archives gouvernementales des Archives nationales du Canada.

Ce numéro du Bulletin est consacré aux travaux entrepris dans le domaine des normes. Plusieurs des projets concernant la description des documents informatisés dans des répertoires ou des catalogues collectifs, ou le transfert de ces documents d'un système à un autre. Ces travaux sont importants en ce qu'ils aident à procurer des données sous une forme qui se prête au transfert et à la préservation, et à fournir à divers groupes de chercheurs de l'information sur les fonds de données.

Format des documents du CULDAT

Comme nous l'avions déjà mentionné dans le dernier numéro du *Bulletin*, le projet de Catalogue collectif canadien de fichiers de données ordinolingues (CULDAT) se poursuit. Au cours de la première phase du projet, une norme provisoire en matière de notice a été conçue. Depuis, on a élaboré et discuté une norme révisée. Le présent article renferme une description de la norme qui sera utilisée pour les notices descriptives. Dès le point de départ, lorsqu'il s'est agi de choisir le contenu et la forme de la description des fichiers ordinolingues du CULDAT, la compatibilité avec les normes existantes de contrôle bibliographique des fichiers informatiques a été un critère important. On a tenté de tenir compte, autant que possible, de ces normes, ainsi que des besoins divergents associés à diverses applications et à divers produits, comme un inventaire des données.

La norme descriptive du CULDAT ne touche que les éléments de données considérés comme les plus utiles à la réalisation du principal objectif du projet, qui est de constituer une source centralisée d'information sur les fichiers informatisés existants au Canada qui aidera les chercheurs à déterminer et à situer les fichiers qui pourraient leur être utiles. Il existe six catégories fondamentales d'éléments de description des fichiers ordinolingues :

1) les éléments bibliographiques, qui permettent d'identifier un fichier;
2) les éléments descriptifs, qui permettent de décrire le contenu d'un fichier;
3) les éléments de classification, qui permettent de grouper les fichiers analogues;
4) les éléments d'ordre technique et physique, qui permettent d'accéder à un fichier;
5) les éléments nécessaires à l'analyse ou à l'utilisation d'un fichier;
6) les éléments d'archivage, qui permettent d'archiver et de tenir un fichier.

Pour les besoins du CULDAT, les deux catégories les plus importantes sont les éléments bibliographiques et descriptifs. L'information d'ordre technique et physique et à l'analyse des fichiers ne rentrent pas dans le plan du catalogue collectif national. Ces données seront plutôt fournies par l'établissement d'archives de données informatiques compétent. Toutefois, les méthodes utilisées pour les études sont décrites, car ces renseignements sont considérés comme de l'information descriptive importante pour celui qui tente de déterminer pour la première fois si un fichier informatique pourrait avoir un intérêt pour lui. La nature même d'un inventaire informatisé rend non nécessaire l'utilisation d'éléments de classification; en effet, le groupement des fichiers informatiques analogues se fait en même temps que la recherche et l'extraction des données nécessaires. On a inclus plusieurs éléments de classification à titre de zones discrètes pour permettre un découpage aisé de la base de données; par exemple, la zone "Type de fichier" permet à l'utilisateur d'extraire tous les fichiers d'une même série chronologique numérique, puis de faire des recherches en se servant d'autres critères.

Voici la liste des éléments de données du CULDAT, classés par catégorie :

Éléments bibliographiques

Chercheur(s) principal(aux) : dans cette zone, on note le nom du (des) principal(aux) responsable(s) de la création du contenu intellectuel du fichier informatique.

Chercheur(s) collectif(s) : lorsqu'aucun enquêteur principal n'est nommé, on inscrit dans cette zone le nom d'au moins une collectivité-auteur responsable en majeure partie de la création du contenu intellectuel du fichier informatique.

Titre : cette zone renferme le titre entier du fichier ordinolingue, ainsi que le sous-titre et (ou) le titre parallèle (titre traduit dans une autre langue).

Variante de titre : dans cette zone est inscrit un titre qui diffère du titre mentionné dans la zone "Titre" et qui est considéré par le catalogueur comme un renseignement additionnel important à des fins d'identification ou comme un point d'accès supplémentaire utile.

Édition : cette zone contient de l'information concernant une édition d'un fichier informatique qui diffère d'autres éditions.

Production : cette zone renferme de l'information sur la date de production du fichier informatique, le nom du producteur et la date de production. Le producteur d'un fichier informatique est la personne ou la collectivité-auteur qui est responsable, administrativement, des mécanismes physiques qui créent le fichier automatisé. La date de production est la date où le fichier est devenu utilisable sous une forme informatisée et à partir de laquelle il a pu être analysé, traité et peut-être même mis à la disposition du public.

Distributeur : le nom du distributeur et le rang attribué par le distributeur au fichier informatique, le cas échéant, figurent dans cette zone. Le distributeur est la personne ou l'organisme chargé par l'enquêteur ou le producteur de faire des copies de fichiers à distribuer.

Générateur : le nom du générateur figure dans cette zone. Le générateur est la personne ou la collectivité-auteur qui est responsable des mécanismes physiques qui mettent un fichier sous une forme informatisée. Il n'a pas d'autre responsabilité que la conversion des données.

Importance du fichier : cette zone nous renseigne sur le nombre d'éléments (c'est-à-dire de fichiers) qui composent un fichier ordinolingue et sur le nombre d'enregistrements logiques que comporte chaque fichier.

Mention de collection : si le fichier ordinolingue fait partie d'une collection de monographies, cette zone renfermera le titre de la collection, ainsi que le volume ou le numéro de cette composante, si les éléments de la collection sont numérotés.

Commanditaire : lorsque le fichier ordinolingue résulte d'un projet financé par un particulier ou un organisme, on note de l'information sur celui-ci dans cette zone.

Éléments descriptifs

Note de dépouillement : cette zone renferme une note structurée qui nous renseigne sur le titre et l'importance des composantes

¹ Dodd, Sue A. *Cataloguing Machine-Readable Data Files: An Interpretive Manual*. Chicago : American Library Association, 1982, 247 p.